A Network Telescope perspective of the Conficker outbreak

Barry Irwin
Security and Networks Research Group (SNRG)
Department of Computer Science, Rhodes University, Grahamstown, South Africa
Email: b.irwin@ru.ac.za

***Abstract*—This paper discusses a dataset of some 16 million packets targeting port 445/tcp collected by a network telescope utilising a /24 netblock in South African IP address space. An initial overview of the collected data is provided. This is followed by a detailed analysis of the packet characteristics observed, including size and TTL. The peculiarities of the observed target selection and the results of the flaw in the Conficker worm's propagation algorithm are presented. An analysis of the 4 million observed source hosts is reported by grouped by both packet counts and the number of distinct hosts per network address block. Address blocks of size /8, 16 and 24 are used for groupings. The localisation, by geographic region and numerical proximity, of high ranking aggregate netblocks is highlighted. The paper concludes with some overall analyses, and consideration of the application of network telescopes to the monitoring of such outbreaks in the future.**

***Index Terms*—Conficker, Zotob, malware, network telescope**

I. Introduction

THIS paper explores the application and value of the use of a network telescope [1], [2] in the tracking and monitoring of a global malware outbreak, over the period from August 2005 to September 2009. During this period, the volume of traffic observed arriving at the research telescope destined for port 445/tcp grew dramatically, particularly over the last 14 months of the dataset, reaching a peak of nearly two orders of magnitude higher than the previously observed traffic baseline. Much of this can be likely attributed to the prevalence of the Conficker worm [3], also known as Kido and DownAdUp [4].

This malware exploits a vulnerability in the Microsoft RPC stack detailed in the Microsoft MS08-067 [5] security bulletin released on 23^{rd} October 2008. The vulnerability exploited is similar to those discovered in July 2003 detailed in MS03-026 [6] and later in MS03-039 [7] — and subsequently exploited by the Blaster and Welchia worms in August of that year [8]. A further vulnerability in the RPC stack was exploited by Sasser in April 2004, taking advantage of the vulnerability disclosed in MS04-011 some seventeen days previously [9]. The problems with the RPC/DCOM stack in Microsoft Windows Family operating systems continued and MS06-40 released in September 2006 [10], patched a further vulnerability that was exploited by various malware such as Mocbot. Given this history of vulnerability, and the widespread adoption of the Windows operating platform and the rapid development of code exploiting these vulnerabilities, researchers were justifiably concerned when the MS08-067 vulnerability was announced. A detailed analysis of the Conficker malware is beyond the scope of this research. For details on the actual origins, and analysis from a payload and reverse engineering perspective, readers are encouraged to consult in particular the work done by SRI [11] and Symantec [12] on reverse engineering and documenting the spread.

This paper presents a discussion of how the spread of this malware was observed from the perspective of the network telescope system system, using data as described in Section II. An overview of the evolution of the worm is presented along with a time-line of the major points in the evolution of this software in Section III. This is shown to match fairly accurately with the observed changes in traffic presented in Section IV. An analysis of the traffic is presented in Section V, with a focus of on the traffic distribution across the target addresses in Section VI. These latter two sections present evidence for the strong likelihood of the majority of the 445/tcp traffic being Conficker related. The paper concludes with a reflection on the application of a network telescope to the monitoring of this kind of event, and the views of traffic as observed by other sensors.

II. Data Source

This research was carried out using TCP/IP packet data collected over a 48 month period period from August 2005 to September 2009, using a /24 IP network address block, within South African address space. The address space was connected to a passive network telescope [13] operated by the researcher [14]. This data includes the period of the outbreak of the Conficker worm in late 2008 on which this paper focusses. What is important to bear in mind when analysing the data collected using the Rhodes University system, is that one of the shortcomings of the current network telescope setup is that only the first packet of the potential TCP 3-way handshake is actually captured. Since the handshake, by design, cannot complete, no data payload can be captured. Due to this limitation it can only be inferred, albeit with a high level of certainty, that the increase in observed traffic is directly related to the Conficker malware. It is believed that the majority of the recorded connection attempts are automated connections from Conficker, but there is certainly a component which is scanning activity from other sources looking for operational targets which may also be vulnerable to the MS08-067 issue.

III. Conficker Evolutionary Time-line

The evolution of the threat posed by the Conficker malware can be traced back to the release of the MS08-067 advisory on 23^{rd} October 2008 as an emergency, out of sequence, patch by Microsoft after exploitation of the vulnerability was observed in the wild. One of the issues to be aware of when analysing Conficker and research around the threats, relates to the two different naming conventions used by Microsoft, and the Conficker Working Group (CWG). The former appears to be in more widespread use. These differences are shown Table I. In this document the Microsoft naming conventions are used. When analysing the traffic, inflexion points can be seen relating to the version changes in the Conficker malware, as seen in Section IV-B. A more detailed timeline of the evolution of this threat is maintained by the CWG[1].

Table I
Conficker Naming

Date	Microsoft	CWG
20 Nov 2008	Conficker.A	Conficker.A
28 Dec 2008	Conficker.B	Conficker.B
20 Feb 2009	Conficker.C	Conficker.B++
4 Mar 2009	Conficker.D	Conficker.C
8 Apr 2009	Conficker.E	-

IV. Telescope traffic observations

Observed network traffic destined to 445/tcp makes for an interesting case study on a number of fronts. Firstly, it is the single most significant contributor to the total, both in terms of the number of packets and source addresses observed. Secondly, it is used by the Microsoft Windows family of operating systems for RPC/DCOM communications, including file sharing, and is usually enabled on such systems. The popularity of the deployment of these systems make this a historically inviting target when vulnerabilities are found, with widespread exploitation. Furthermore, this port is generally firewalled by most organisations, and often by home users as well, although usually only for inbound traffic.

A. Overview

Traffic destined to port 445/tcp as a whole, can be seen to be fairly persistent over the entire duration of the network telescope observation, being observed on all but one of the 1 429 days having data (and 98.1% of hourly observations) within the dataset. Over the period it was consistently ranked in the top ten ports observed, by both month and year. During the observation period, packet counts for traffic destined to port 445/tcp was the top ranked in 10 of the 17 quarters under study, with its lowest positions being 4th in Q1 2007 and Q4 2008. Figure 1 shows the prevalence of this traffic over the observation period. Data shown in this figure reflects only that TCP traffic destined to 445/tcp that has the SYN flag set, and hence can be considered 'active' (in terms of this traffic could potentially elicit a valid response from a target). This traffic also accounted for 41.4% of the traffic overall,

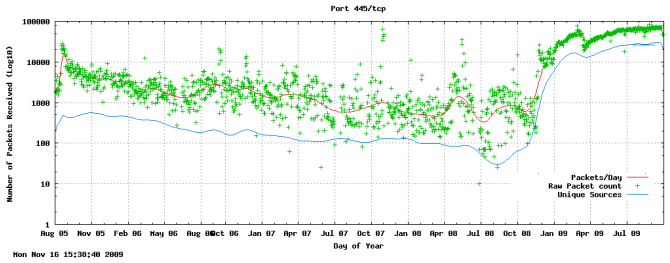

Figure 1. TCP packet received on port 445 by day.

consisting of some 16 893 920 packets. The rapid increase in traffic from approximately 1 000 packets/day (ppd) in October 2008 through to nearly 100 000 ppd by the end of September 2009 can be clearly seen. The benefit of the long collection baseline was realised in this comparison, in terms of being able to quantify just how large the increase in traffic was.

The spike in observed activity the in early portion of Figure 1 is most probably attributable to the Zotob worm [15] exploiting a vulnerability disclosed on 9^{th} August 2005 in MS05-039 [16], [17], either from the worm itself or related scanning in response to this event by individuals looking for vulnerable hosts. Traffic levels had, however, decreased and largely normalised by November 2005 and continued to drop though to mid October 2008. This gradual decrease is likely due to the increased uptake of automated patching of systems though the Microsoft Windows Update mechanism, the release of Service Packs for Windows XP (SP3 - April 2008) and Windows Vista (SP1 - March 2008, SP2 - April 2009) resulting in the remediation of vulnerabilities in the RPC service. More significantly the lack of any significant vulnerabilities affecting this protocol during the observation period, would have reduced the incidence of scanning. The rapid increase in traffic observed from October 2008 onwards can be attributed to activities surrounding the exploitation of the MS08-067 vulnerability in Microsoft Windows operating systems, most notably by the Conficker worm. The remainder of this paper focuses on activity observed from October 2008 onwards.

B. Conficker Related

The Conficker worm was first observed on the $20^{th}/21^{st}$ of November 2008 (depending on time zone), and after almost a year, over 7 million infected nodes were observed as still infected[2]. After the 20^{th} of November, traffic destined to 445/tcp constituted 70% of traffic observed. Over the entire observation period, 4 002 119 unique hosts (86% of the total) were observed sending packets to a destination port of 445/tcp. Of these addresses, 95% were observed after the 20^{th} of November 2008, with only 5 544 (0.14%) having been observed prior to the 1^{st} of November 2008. Only 0.3% of the IP addresses identified as having targeted 445/tcp had sent any traffic at all, prior to the beginning of November 2008.

This is not the first work to have been done on looking at Conficker from the perspective of a network telescope,

[1]http://www.confickerworkinggroup.org/wiki/pmwiki.php/ANY/Timeline

[2]http://www.shadowserver.org/wiki/pmwiki.php/Stats/Conficker

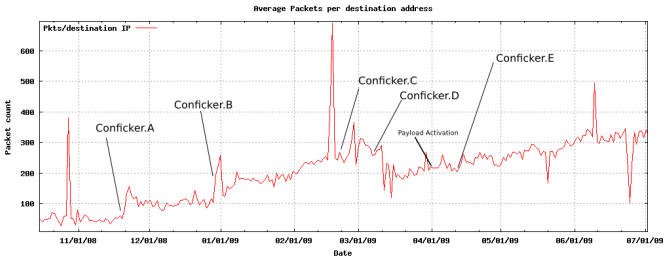

Figure 2. Traffic November 2008 — July 2009

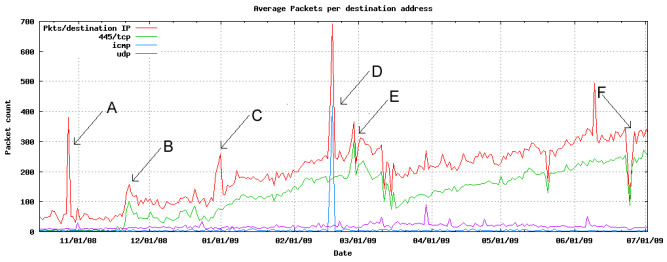

Figure 3. Protocol breakdown (November 2008 — July 2009)

with some detailed analysis having been performed by CAIDA researchers in [18], and the subsequent release of a portion of their *3 Days of Conficker* dataset [19]. What is novel in this paper is the fine level of detail at which it has been performed along with the use of data from a single /24 network telescope, rather than the aggregated level previously reported utilising the CAIDA /8 telescope, and a further dataset gathered on a /16 netblock. The discussion below focuses on the traffic considered Conficker related from mid October 2008 through to the end of the dataset at the end of September 2009, in effect covering nearly a year of activity related to the MS08-067 vulnerability.

An overview of the total traffic observed by the telescope system is shown in Figure 2 as the calculated average number of packets received per IP addresses in the monitored range. A number of distinct spikes in the traffic are noticeable, along with the general increase in traffic over time. The increase is, however, not nearly as rapid as that observable in the latter part of Figure 1. Particularly notable events are the large spike on 28^{th} October 2008, followed by a rapid climb on the 21^{st} November 2008. A second rapid increase can be seen on 1^{st} January 2009, with a consistent increase in traffic rates observed though to mid February, and a large increase in activity on the 28^{th}. This is followed by a sharp drop-off mid March and a small spike prior to 1^{st} April. From this point the traffic continues to increase, other than two dips which were caused by network outages. On initial observation, these periods seem to coincide to those outlined in the evolution timeline in Table I.

Looking a little deeper, and analysing the composition of the traffic down by protocol, one can see that the spikes observed cannot be correlated directly to activity on port 445/tcp. This detailed breakdown of the same dataset and time period as previously shown in Figure 2 can be seen in Figure 3. In the detailed plot, ICMP and UDP traffic have been shown along with the contribution made by traffic destined to 445/tcp on the sensor network. Several points in Figure 3 are worth highlighting:

- Although the spike shown at point A ties in with the release of the MS08-067 security bulletin, it is not related to it, but rather is the result of a burst of classic back-scatter packets originated from a unix system located in Jordan.
- The 'birth' of Conficker on the 21^{st} of November 2008 (point B) can be seen by the sharp rise in 445/tcp traffic as a portion of the whole.
- Conficker.B was released on the 28^{th} of December 2008 and although there is a spike in total traffic (point C) this was not due to Conficker, but a reflected back-scatter SYN-ACK packets from a web server located in Costa Rica.
- Point D indicates a further anomaly in the traffic pattern, with the spike caused by a flood of 159 000 ICMP TTL expired messages, received from a host in China on the 17^{th} and 18^{th} of February 2009.
- The spike in late February 2009 (point E) can be attributed to the release of Conficker.C on the 20th.
- Point F is worth noting in that the drop in recorded traffic was due to a series of extended network outages adversely affecting Rhodes University's upstream Internet connection over the period 24^{th}-26^{th} June 2009.

C. Conficker Outbreak

Analysis of the data in the first few days of the Conficker outbreak revealed some interesting trends. The first of these is illustrated in Figure 4, which plots data relating to traffic received on 445/tcp during the period $21^{st} - 24^{th}$ November 2008. Times noted in the Figure are SAST (GMT+2:00). What is immediately noticeable is that while the packets follow a rough circadian rhythm, this trend is even more noticeable when the number of distinct sources for each hour interval are plotted. Similar results were found with the data processed in the CAIDA Conficker report [18]. Considering the 24 hour period from midnight on the 21^{st} November, the number of observed sources per hour can be seen to climb rapidly, from fewer than ten at 05h00 to over 250 by midnight the following day. What is interesting is that there is a large increase in packets observed around 06h00, yet only twenty source hosts had been observed at this stage. From this point the packet rate per hour dropped dramatically and the host count started to climb — in essence, more hosts were sending relatively fewer datagrams. By 16h00 traffic had reached a low point and subsequently started to increase again, with a rapid growth in the number of sources observed; a high sustained rate being maintained for nearly ten hours before dropping back. This is pattern can be seen to be repeated over the next two days.

An interesting anomaly was found in the data for 445/tcp, with a significant spike in scanning activity detected over a period from late on the 30^{th} September, through to the evening of the 1^{st} of October 2008. This increase in scanning was particularly noticeable, due to there being almost no traffic targeting 445/tcp in the days leading up to this, and very little

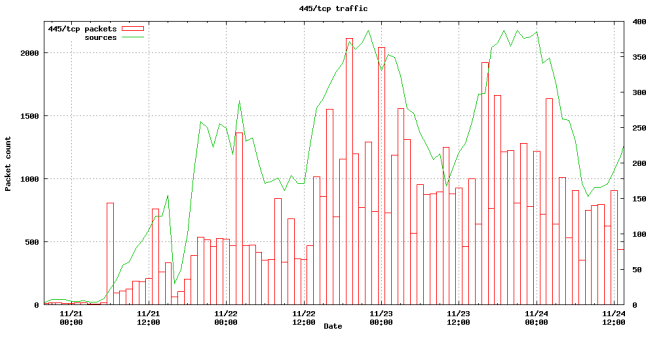

Figure 4. Early days of Conficker: $21^{st} - 24^{th}$ November 2008

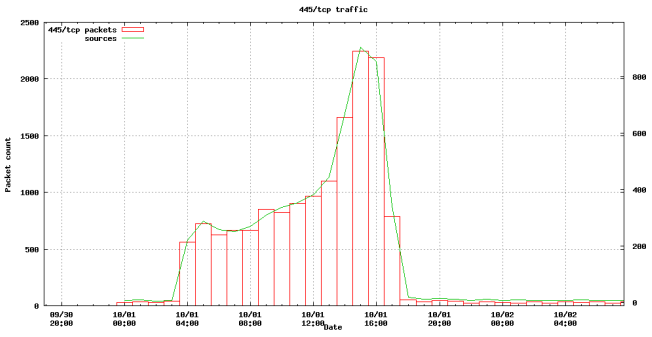

Figure 5. Early reconnaissance: 30^{th} September — 2^{nd} October 2008

afterwards. A plot of the relevant traffic is presented in Figure 5. The traffic is also seen to originate from a relatively high number of sources, with 3 476 IP addresses being logged between 03h00 and 18h00 on the 1^{st} of October, having sent some 14 808 packets targeting 445/tcp. Top geopolitical sources as determined by geolocation tools were Brazil (BR), France (FR), the USA (US), Japan (JP) and the Philippines (PH). Together these accounting for nearly half (46%) of the sources during this period, a summary of which is shown in Table II. Sources were observed from 100 countries, although only 30 of these had more than twenty sources.

Without payloads, it is impossible to determine the exact nature of this traffic, or the means of its generation. It could plausibly be activity from the Gimmiv trojan, although it was noted that there were problems with its replication mechanisms. Another possibility is that it could be some custom malware utilising the Chinese exploit kit based on the Gimmiv trojan. It is nonetheless interesting to note that there is a fairly even distribution across the entire monitored IP range, in contrast to what is seen in Section VI. The majority of the sources (99%) scanned less than ten target addresses, with 1832 (52%) only probing one host. Only three hosts scanned more than 100 addresses, while the majority of sources sent two packets which were in relatively quick succession, before disappearing.

It is the researcher's hypothesis that there is a strong likelihood of this having been a distributed scanning attempt, with multiple sources scanning to look for vulnerable hosts possibly for further targeted exploitation, or as a means of seeding initial distribution points for later malware release. The fact that such a high number of hosts only probed a single target points to a well co-ordinated, distributed scan, or these addresses possibly being used as a decoy scan. The three top hosts were determined to be in Taiwan (TW) and the USA, and are most likely the real hosts. This is supported by the fact that hosts that are geolocated as originating from the Philippines (PH), Croatia (HR), Turkey (TR) and Austria (AT) all have TTL values above 240. This is highly unlikely given the university's Internet connectivity which had, at the time of collection, at least ten hops to get to international peering points in London. Further examination of the TTL values shows that a significant number of the hosts have the same TTL values despite being geolocated to vastly different parts of the globe, further strengthening the likelihood of packet forgery. The fact that only two packets are sent is also interesting as generally most TCP/IP stacks send three SYN connection attempts before timing out. This could point to the fact that custom code was being used with a short time-out, or that packets were being constructed using 'raw' sockets.

Table II
TOP 5 COUNTRIES - 1^{st} OCTOBER 2009

Rank	CC	$Count_{Packets}$	$Count_{Source}$	$Count_{Dest}$
1	BR	2 203	625	249
2	FR	1 551	388	231
3	US	1 877	279	249
4	JP	654	167	177
5	PH	573	155	157
Total		6 858	1 614	

V. PACKET DATA

This section evaluates aspects of the packets received on 445/tcp by the network telescope, considering the observed TTL, packet structure, packet retransmission, and source operating system.

A. Time to Live

An analysis of the TTL values recorded for all incoming traffic destined to 445/tcp showed a very narrow banding where it was observed that the values were, with few exceptions, between 50 and 100. This range covers default TTL settings for both Windows and unix platforms, having default base TTL values of 128 and 64 respectively. This banding is further evident when plotted against packet counts for TTL values received overall as presented in Figure 6. In this Figure the TTL values for 445/tcp packets, can be seen to be largely grouped in the 96-128 range, with very few packets recorded in the 32-64 and 224-255 ranges. This again provides strong evidence towards Microsoft Windows platforms being the origin of the packets. This was confirmed by the passive operating system fingerprinting that was performed. This would in turn lend weight to the supposition that when considering the number of distinct sources, the packets observed were actually generated by the automated scanning modes of the Conficker worm.

B. Packet Structure

Based on an analysis of of the datagrams received, the majority were found to be of 62 bytes in size. A summary

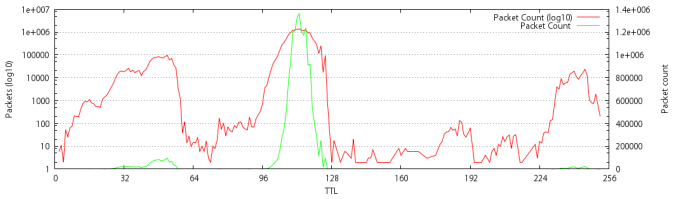

Figure 6. Packet counts by TTL for 445/tcp

Table III
SIZE OF PACKETS TARGETING 445/TCP

Size	$Count_{Packet}$	Packet %	$Count_{Source}$	Source %
62	12 207 669	86.18	3 248 492	85.16
66	1 281 603	9.04	443 325	11.62
78	423 043	2.98	153 076	4.01
60	150 549	1.06	25 211	0.66
74	100 043	0.70	31 278	0.81
	Total	99.96		102.26[a]

$N_{Packets} = 14\,165\,097 \quad N_{Sources} = 3\,814\,447$
[a] This is > 100% as hosts may have sent different size packets

of packet sizes is given in Table III. In this table, it can be seen that more than 85% of both the hosts and packets match this sizing. This is reached by having a TCP packet with no data and encapsulated inside an IP and finally an Ethernet datagram. In order to reach the value of 62 bytes, rather than the default of 60, TCP options are set. The combination found set most often was "MSS:1460, NOP NOP TCP SACK:TRUE", which accounted for 8 bytes. These options enabled the Selective Acknowledgement (SACK) on the connection being established, along with a maximum sent size of 1 460 bytes. These settings were found to match captured Conficker propagation traffic. Thus there is a fairly high probability that the TCP SYN packets being sent to addresses on the telescope actually originated from the Conficker Malware.

This provides an example of how, despite being somewhat handicapped by the lack of payloads in a network telescope dataset, comparative data from honey-net or other systems with higher levels of interaction can be used to augment the analysis process. While absolute certainty is not possible without payload analysis, researchers can attain a high level of confidence in their analyses.

C. Transmission

A further interesting characteristic observed in the 445/tcp traffic after the advent of Conficker, is that it has a very noticeable signature 'on the wire' in terms of the way connection attempts are made. Most operating system TCP/IP stacks will send at least three TCP SYN packets in an attempt to establish a connection. By contrast, in the majority of the 445/tcp traffic received after the 20^{th} of November 2008, one observes only two connection attempts. An example of this is shown in Figure 7, where source addresses (in the 4^{th} column) make two connection attempts approximately three seconds apart. Similar behaviour has been observed by [18]. This was further validated by the researcher with captures of live propagation traffic obtained from hosts with confirmed Conficker infections.

Table IV
TRAFFIC TO 445/TCP BY ATTRIBUTED OPERATING SYSTEM

Rank	Protocol	Number	%
1	Windows	16336052	99.671
2	Proxyblocker	19401	0.118
3	MacOS	10066	0.062
4	FreeBSD	7114	0.043
5	Linux	4361	0.026
6	NetBSD	3981	0.024
7	Cisco	3230	0.019
8	Solaris	1910	0.011
9	Checkpoint	1343	0.008
10	NMAP	1258	0.007
		Total	99.992

N=16 389 887
% of Packets attributable to an IP address with an identified OS
N is calculated as packets received after 2008-11-20 00:00 GMT+2

Considering the total number of sources observed and the total number of packets targeting 445/tcp after November 20^{th}, these are in a ratio of approximately 1:4, indicating that on average most sources scanned two hosts at most. This is discussed further in Section VI, where an analysis of the target addressing is provided.

D. Operating System Fingerprinting

Operating System attribution was performed using p0f[3] - a passive operating system fingerprinting tool developed by Michal Zalewski. While it is recognised that this method is not flawless, and may be skewed by the use of NAT and dynamic address allocation, it nevertheless provides a useful measure. The results are presented in Table IV. Microsoft Windows family platforms accounted for 99.7% of the sources that could be attributed, which is is unsurprisingly given the facts that the Conficker malware targets these platforms, and the TTL data as seen in Section V-A.

VI. TARGET ANALYSIS

Changing focus away from the sources of the traffic to the addresses being targeted in the network telescope address space, a very uneven distribution pattern is observed. The lower half of the monitored space, i.e. 196.x.x.0/25, is targeted substantially more than the upper half (196.x.x.128/25). Particularly heavily targeted is 196.x.x.1, closely followed by other addresses in the lower 16. The first eight addresses in the address block all received more than 100 000 distinct sources. This bias is shown in Figure 8, which considers the number of distinct sources rather than packets observed for each IP address in the monitored range.

The strong bias towards the lower portion of the address space can be seen clearly. Notably, the last address in the monitored range (196.x.x.255) received a much higher coverage than other IPs in the upper /25 portion. The reason for this bias is most likely a naive scanning optimisation, which attempts to probe one or more addresses on a range, and if no response is received, moves onto another range. The probes to the last address on the range may serve a similar purpose. By

[3] http://lcamtuf.coredump.cx/p0f.shtml

```
1 01:16:21.685360 IP 190.50.x.80.2725 > 196.x.y.3.445: S 2062323770:2062323770(0)
2 01:16:23.509228 IP  77.28.x.55.4853 > 196.x.y.34.445: S 1323192692:1323192692(0)
3 01:16:24.677814 IP 190.50.x.80.2725 > 196.x.y.3.445: S 2062323770:2062323770(0
4 01:16:26.514630 IP  77.28.x.55.4853 > 196.x.y.34.445: S 1323192692:1323192692(0)
5 01:16:27.693010 IP  79.0.x.248.1731 > 196.x.y.18.445: S 1786561877:1786561877(0)
6 01:16:29.808481 IP 189.101.x.133.2499 > 196.x.y.3.445: S 3114908412:3114908412(0)
7 01:16:30.696890 IP  79.0.x.248.1731 > 196.x.y.18.445: S 1786561877:1786561877(0)
8 01:16:32.751635 IP 189.101.x.133.2499 > 196.x.y.3.445: S 3114908412:3114908412(0)
```

Figure 7. Examples of the two packet repetitions

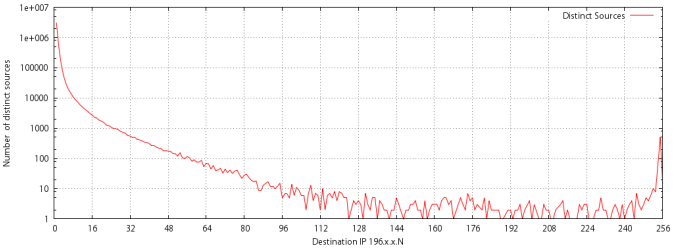

Figure 8. 445/TCP traffic: Distinct Sources per sensor IP

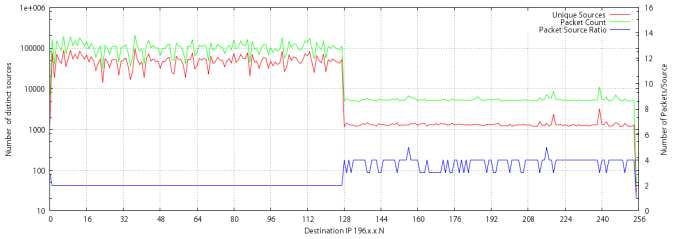

Figure 9. 445/TCP total traffic received per IP in the sensor network

convention default gateways on most IP networks either make use of the first or last address in a given subnet.

Considering the packet counts received for each address, presents a markedly different picture, as shown in Figure 9. A significant change in the levels can be seen occurring at 196.x.x.128, with a drop in recorded traffic of nearly two orders of magnitude. The Figure contains the number of sources, and the packet count observed for each address. A third value is plotted on the right *y-axis*, being a ratio of the number of packet to distinct sources. This value also makes a change as the target address moves into the upper half of the range, doubling from just over 2 to 4. The dip at the beginning of the graph is due to some traffic having been recorded to 196.x.x.0, which is not normally targeted as it it would be regarded as the 'network address', rather than a host address.

This observed bias towards the lower 128 hosts in the network is due to a bug in the scanning algorithm implemented by Conficker [20], [21]. Due to the way the pseudo random number generator works, a 15-bit value is generated and is then used in a 16-bit field, resulting in the most significant bit of the 2^{nd} and 4^{th} octets (bytes) in an IPv4 address to being zero; in effect limiting these values to the range 0-127. The side effect of this is that it significantly reduces the portion of the Internet that can possibly be scanned [18], [22], limiting it to only 25% of total address space — although it will attempt to scan all the /8 blocks and half of all /16 blocks. Taking this into account, one can use the traffic to the upper 128 address to quantify the other scanning activity present on the port that is not attributable to the Conficker Worm, but rather other sources.

Looking at the sources of the traffic directed to the lower half of the monitored range, 26% (1 049 562) had a 2^{nd} octet greater than 127, and 46% (1 842 784) with the last octet having a value greater than 127. Common to both these groupings were 515 834 hosts. What is interesting is that these these could not have been infected by direct scanning, although given dynamic addressing they could have had other addresses previously, or have changed networks (as is common with mobile systems). Infection is still possible though other measures such as via the Windows 'autorun' mechanism on removable media, which was used by the earlier Conficker variants. These sources are analysed further in Section VII.

It is worth noting that the differences in traffic observed between the upper and lower ranges is substantially more than the three times differential described by [22] and attributed to this activity. This bug most likely accounts for the fact that the a second telescope operated by the researcher recorded minimal traffic destined to 445/tcp, since both its 2^{nd} and 4^{th} octets are greater than 127. In addition to this bug in the way random IPs are generated, IP ranges known to belonging to security research firms, and Reserved IPv4 Address space are also avoided by the malware. This is an interesting case which shows the value in having distributed IP address space for monitoring global trends. It is also an argument for chunks of contiguous space rather than smaller fragmented blocks as advocated with the use of greynets [1].

VII. Source Analysis

Analysing the source address data for the Conficker provides an interesting insight, especially when comparing the top sources ranked by packet count and host count. These are presented in Tables V and VI respectively, and discussed in the sections below. The value in discriminating between these two means of ranking data is that the first quantified the volume, and to some extent the persistence of the infection. In contrast, ranking and subsequent analysis by the number of distinct sources observed provides a means for assessing how widespread the infection and related scanning activity was.

A. Packet Count

Considering the rankings by /8 netblock, 196.0.0.0/8, managed by AfriNIC, can be seen to have the highest packet count by a significant margin. It is worth taking note that a single host within this block accounted for 23% of the packet count in this netblock. Looking at the remainder of the top

ten, over 38% of the total packet count can be observed. The numerical sequencing, of the top ten /8 netblocks have very close numerical groupings. The allocations within 189.0.0.0/8, 190.0.0.0/8 and 201.0.0.0./8 are controlled by LACNIC, with the remainder of the top ten being under the control of RIPE. Three adjacent groups of netblocks are observed. This is most likely due to the scanning and propagation mechanisms used by the Conficker malware.

This numerical closeness is also evident when considering the /16 rankings, with two contiguous address blocks in positions one and two. This can also be seen with the two blocks in 93.80.0.0/15. In the /24 rankings four contiguous blocks can be observed in 196.20.164.0/24, contributing to the high rankings of both the 196.20.0.0/16 and subsequently 196.0.0.0/8 netblocks. However, even combined, these four blocks comprising 1024 addresses account for less than a third of the volume the top observed host. The traffic attributable to individual hosts shows a rather dramatic decrease, with the top ranked host of 196.21.218.x accounting for more than two and a half-times the sum of the remainder of the top ten hosts. A second host from the same /24 netblock appears in 6^{th} position.

A comparison of the percentage contribution for each of the netblock aggregation levels decreases sharply from the level covered by the /8 netblocks. However there is relatively little difference between the /24 and /32 levels, largely due to the contribution by the hosts in 196.21.218.0/24. It is still significant that the top host rankings accounted for 2.5% of the total, despite there being over 3.8 million hosts observed, even more so the percentage of traffic attributable to the top ranked hosts. The nearly 80% decrease in composition moving from /8 to /16 is an indicator of how widely spread the scanning activity was, although the proportion covered (7.9%) is less than the 12% observed in the overall dataset, with the values for /24 and host level proportions being much closer.

B. Source Count

Rankings of the number of distinct sources observed as origins for traffic destined to 445/tcp, aggregated of by network block is presented in Table VI. This presents a somewhat different picture to that discussed in the previous section. This ranking provides an indication of how widespread the activity was within a particular netblock, and can be used as a means of determining infection rates. These are not absolute values, and may well be influenced by the use of NAT and dynamic addressing of endpoints, or even a combination of these — as is particularly common in modern broadband networks.

At the /8 aggregation level, the first eight ranked netblocks also occur in the top ten by packet count, with 196.0.0.0/8 and 93.0.0.0/8 being omitted. These placed 54^{th} (with 14 914 sources) and 72^{nd} (4 268 sources) respectively. The top four aggregated blocks had a fairly equitable portion of the total hosts served, at just over 5% each. The top ten accounted for over 40% of the total hosts. Again sequential netblock are seen to occur indicating a strongly likelihood of propagation activities favouring 'near' or 'local' addresses, although this could have been influenced by some of the other propagation mechanisms used. This sequential locality is repeated again with the /16 aggregation. A strong geopolitical bias can be seen at this level, with five of the top 6 ranks netblocks being under the control of Corbina Telecom (AS8402), head-quartered in Moscow, Russia.

Finally when considering the top ten /24 netblocks, the top source network (196.1.232.0/24) originates from the Sudan (SD) operated by SudaTel (AS15706). This is significant as 233 individual IP addresses have been observed out of a possible 254 operable host addresses, providing a coverage factor of over 90%. Further sequential address blocks can be seen with 79.114.134.0/23, and the aggregate block of 89.179.104.0/22. Other /24 blocks in the top also exhibit both numerical and topological (being part of the same ASN in global routing tables) closeness. Of interest is that despite 89.179.0.0/16 containing five of the top ten /24 netblocks, is is not in the top ten when aggregation is performed by /16, appearing only in 39^{th} place in these rankings.

VIII. Future Work

The dataset used in the research can still be further analysed, particularly from the point of an extended temporal or geopolitical analysis, such as that performed in [14]. Since the Conficker outbreak, there has not been another significant Internet scale event. As such, further exploration of the dataset on which this work is based, and other subsequently collected datasets may provide better insight into the spread of malware and related malicious activity on a global scale.

IX. Conclusion

This focused analysis of traffic destined to 445/tcp has covered two distinct global malware threats — that of Zotob in August 2005, and Conficker in November 2008. In the intervening period traffic levels remained consistent, and can be attributed to remnants of the Zotob malware and similar other software, and scanning by individuals for hosts having services on 445/tcp exposed to the Internet at large in order to potentially exploit their vulnerability. Over the period of the Dataset, and particularly in the last 14 months, traffic destined to 445/tcp made a significant contribution to the whole. Given this, it is important to investigate the nature and origins of the datagrams.

While the analysis carried out in this paper is by no means complete, it provides an example of the kind of focused analysis that can be done with a network telescope. The evolution of the Conficker worm is plotted. The problem with the random scanning and propagation algorithm identified in the reverse engineering of the malware can be clearly observed, and this is seen to be a plausible explanation for the significant difference in traffic observed by the researcher between the dataset being considered for this work, and others utilising different address space. Furthermore, the work presented shows how a network telescope can be used to track the spread and distribution dynamics of widespread internet worms in the future.

Table V
TOP 445/TCP ORIGINS BY PACKET COUNT

Rank	/8	count	%	/16	count	%	/24	count	%	/32	count	%
1	196.0.0.0	1 118 121	7.921	196.21.0.0	287 270	2.035	196.21.218.0	266 790	1.890	196.21.218.x	258 615	1.832
2	189.0.0.0	600 718	4.255	196.20.0.0	234 092	1.658	196.20.164.0	22 734	0.161	196.20.17.x	21 611	0.153
3	190.0.0.0	586 827	4.157	78.106.0.0	91 845	0.650	196.20.17.0	21 655	0.153	196.14.169.x	21 576	0.152
4	92.0.0.0	565 721	4.007	93.80.0.0	91 248	0.646	196.14.169.0	21 576	0.152	196.20.13.x	15 992	0.113
5	95.0.0.0	539 401	3.821	93.81.0.0	83 169	0.589	196.20.165.0	20 762	0.147	196.21.125.x	8 848	0.062
6	89.0.0.0	494 255	3.501	95.28.0.0	76 423	0.541	196.38.187.0	20 077	0.142	196.21.218.x	8 175	0.057
7	78.0.0.0	463 784	3.285	89.178.0.0	75 996	0.538	196.20.167.0	18 283	0.129	59.162.166.x	6 549	0.046
8	79.0.0.0	387 708	2.746	95.24.0.0	68 342	0.484	196.20.166.0	17 758	0.125	196.32.152.x	6 302	0.044
9	93.0.0.0	367 343	2.602	190.51.0.0	54 041	0.382	196.20.13.0	15 992	0.113	196.15.239.x	5 958	0.042
10	201.0.0.0	353 829	2.506	196.205.0.0	53 587	0.379	196.20.140.0	14 668	0.103	196.34.217.x	5 841	0.041
Total			38.801			7.902			3.115			2.542

$$N_{Packets} = 14\,115\,791$$

Table VI
TOP ORIGIN NETBLOCKS FOR 445/TCP BY SOURCE COUNT

Rank	/8	Sources	%	/16	Sources	%	%	/24	Sources	%
1	189.0.0.0	209 921	5.511	78.106.0.0	30 407	0.798	46.397	196.1.232.0	233	91.015
2	92.0.0.0	204 970	5.381	93.80.0.0	29 643	0.778	45.231	79.114.134.0	229	89.453
3	190.0.0.0	191 752	5.034	93.81.0.0	25 766	0.676	39.315	79.114.135.0	226	88.281
4	95.0.0.0	190 630	5.004	95.28.0.0	25 703	0.674	39.219	89.179.78.0	224	87.500
5	79.0.0.0	151 942	3.988	89.178.0.0	25 677	0.674	39.179	89.179.104.0	222	86.718
6	78.0.0.0	143 154	3.758	95.24.0.0	23 763	0.623	36.259	89.179.105.0	215	83.984
7	201.0.0.0	113 530	2.980	190.51.0.0	19 982	0.524	30.490	89.179.106.0	213	83.203
8	89.0.0.0	113 106	2.969	77.28.0.0	17 173	0.450	26.203	89.179.107.0	211	82.421
9	59.0.0.0	111 005	2.914	59.93.0.0	15 446	0.405	23.568	93.81.128.0	209	81.640
10	87.0.0.0	110 858	2.910	77.29.0.0	14 024	0.368	21.398	93.81.132.0	209	81.640
		Total	40.449		Total	5.970				

$$N_{SourceIP} = 3\,809\,104 \quad N_{/8} = 194 \quad N_{/16} = 14\,896 \quad N_{/24} = 495\,602$$

ACKNOWLEDGEMENTS

This work was performed in and funded by the Telkom Centre of Excellence in Distributed Multimedia at Rhodes University. Funding was also received from the National Research Foundation Thutuka Program Grant number 69018 and the Rhodes University Joint Research Committee (JRC).

REFERENCES

[1] F. Baker, W. Harrop, and G. Armitage, "IPv4 and IPv6 Greynets." RFC 6018 (Informational), Sept. 2010.

[2] W. Harrop and G. Armitage, "Defining and evaluating greynets (sparse darknets)," in *LCN '05: Proceedings of the The IEEE Conference on Local Computer Networks 30th Anniversary*, (Washington, DC, USA), pp. 344–350, IEEE Computer Society, 2005.

[3] Microsoft, "Virus alert about the Win32/Conficker worm (KB962007)." Online, August 18 2008. Last Review: December 1, 2010 - Revision: 10.0.

[4] Microsoft, "Win32/conficker." Online, 8 Jan 2009. Updated: Nov 10, 2010.

[5] Microsoft, "MS08-067 : Vulnerability in Server Service Could Allow Remote Code Execution (KB958644)," tech. rep., Microsoft, Oct 23 2008.

[6] Microsoft, "MS03-026 : Buffer Overrun In RPC Interface Could Allow Code Execution (KB823980)," tech. rep., Microsoft, July 16 2003. Originally posted: July 16, 2003 Revised: September 10, 2003.

[7] Microsoft, "MS03-039 : Buffer Overrun In RPCSS Service Could Allow Code Execution (KB824146)," tech. rep., Microsoft, September 10 2003.

[8] Microsoft, "Virus alert about the Nachi worm (KB826234)." Online, August 18 2003.

[9] Microsoft, "MS04-011: Security Update for Microsoft Windows (KB835732)," tech. rep., Microsoft, April 13 2004. Updated: August 10, 2004.

[10] Microsoft, "MS06-040 : Vulnerability in Server Service Could Allow Remote Code Execution (KB921883)," tech. rep., Microsoft, September 12 2006.

[11] P. Porras, H. Saidi, and V. Yegneswaran, "An analysis of conficker's logic and rendezvous points," tech. rep., SRI International, 4 February 2009. Last Update 19 March 2009.

[12] B. Nahorney, "The downadup codex." Online, March 2009.

[13] D. Moore, C. Shannon, G. M. Voelker, and S. Savage, "Network telescopes," tech. rep., CAIDA, 2004.

[14] B. Irwin, *A framework for the application of network telescope sensors in a global IP network.* PhD thesis, Rhodes University, Grahamstown, South Africa, 2011.

[15] B. Schneier, "The Zotob Storm," *IEEE Security and Privacy*, vol. 3, pp. 96–, November 2005.

[16] Microsoft, "Microsoft Security Bulletin MS02-039: Vulnerability in Plug and Play Could Allow Remote Code Execution and Elevation of Privilege (899588)." Online, August 9 2005.

[17] D. White, "MS05-039 and the Zotob summary." Online, 18 August 2005. Last accessed 2010-12-01.

[18] E. Aben, "Conficker/Conflicker/Downadup as seen from the UCSD Network Telescope." Online, CAIDA Network Telescope Project - Backscatter, February 2009.

[19] P. Hick, E. Aben, D. Andersen, and K. Claffy, "The CAIDA UCSD Network Telescope "Three Days Of Conficker" (collection)." Online, CAIDA Network Telescope Project - Backscatter, 2009. Support for the UCSD Network Telescope "Three Days Of Conficker" Dataset and the UCSD Network Telescope are provided by Cisco Systems, Limelight Networks, the US Department of Homeland Security, the National Science Foundation, and CAIDA Members.

[20] M. Richard and M. Ligh, "Making fun of your malware." Conference Presentation Defcon 17, Las Vegas USA, August 2009.

[21] Carnivore.IT, "Conficker does not like me?." Online Blog, 3 November 2009. Accessed 21 November 2010.

[22] E. Wustrow, M. Karir, M. Bailey, F. Jahanian, and G. Huston, "Internet background radiation revisited," in *Proceedings of the 10th annual conference on Internet measurement*, IMC '10, (New York, NY, USA), pp. 62–74, ACM, 2010.